

Oliver Kiechle

Noisy digitalia – Kommunikationslärm in Born Digital-Quellen am Beispiel von Spam im Usenet

Abstract: Historical work that is based on source material from the field of digitalia, thus meaning Born-digital sources, often meets a phenomenon, that is so far unknown to historians in such an extent: Communication noise. It can be found as unwanted advertisement, off-topic comments and even blunt offences. Taking the example of Spam messages in the Usenet of the 1980s and 1990s the article will show how communication nuisance developed during early internet communication and what is its part in historical tradition. With Flaming and Social bots two other manifestations of Noisy digitalia will be considered. Based on that it will be discussed, how historical work can deal with noise in digitalia both in a conceptional and a technical way as well.

Keywords: digitalia, communication noise, social media, Usenet, spam, web history

Zusammenfassung: Bei der geschichtswissenschaftlichen Arbeit mit Quellenbeständen aus dem Bereich der *digitalia*, also genuin digital entstandenem Material, macht sich vor allem im Bereich der Social Media-Quellen ein für die historische Zunft in diesem Ausmaß ungewohntes Phänomen bemerkbar: Kommunikationslärm. Dieser *Noise* kann in Form von unerwünschter Werbung, unpassenden Kommentaren bis hin zu offenen Beleidigungen auftreten und vor allem die automatisierte Analyse größerer Datenbestände beeinflussen.

Im Artikel wird am Beispiel von Spam-Nachrichten im Usenet der 1980er und 1990er Jahre beschrieben, wie sich Lärm in der frühen Internet-Kommunikation entwickelt hat und welchen Anteil er in der Überlieferung einnimmt. Ergänzend werden mit *Flaming* und *Social Bots* zwei weitere Spielarten der *Noisy digitalia* beschrieben. Davon ausgehend wird erörtert, wie mit diesen Störgeräuschen im Rahmen einer historischen Beschäftigung mit *digitalia* sowohl konzeptionell als auch technisch umgegangen werden kann.

Schlagwörter: Digitalia, Communication Noise, Social Media, Usenet, Spam, Web history

Der digitale Austausch ist oft mit inhaltlich unpassenden Einstreuungen oder von externen Akteur:innen platziertem Content durchsetzt, der bestenfalls lästig, schlimmstenfalls sogar destruktiv sein kann. Dieser Kommunikationslärm ist zu einem selbstverständlichen Teil der digitalen Welt avanciert und nimmt dort einen unübersehbaren Stellenwert ein. So kann beispielsweise mehr als die Hälfte der verschickten E-Mails im Jahr 2019 als Spam-Nachrichten klassifiziert werden.¹ Diese Störgeräusche – seien es Werbeanzeigen auf Webseiten, invektive Kommentare in Foren und Chats oder eben die allgegenwärtigen Spam-Nachrichten in Mails oder Newsgroups – beeinflussen nicht nur unsere alltägliche Kommunikation, sondern machen sich ebenfalls bei der historischen Analyse entsprechender archivierter Quellen bemerkbar. Im Folgenden soll deshalb zunächst eine Einordnung von genuin digitalen Materialien im Hinblick auf ihren historischen Kontext gegeben werden. Davon ausgehend werden anhand des Beispiels von Spam-Nachrichten im Usenet der 1990er Jahre verschiedene Aspekte des Auftretens von *Noisy digitalia* erörtert. Danach soll der Blick kurz auf weitere Störfeuer in digitalem Quellenmaterial gerichtet und abschließend Auswirkungen auf die historische Analyse thematisiert werden.

1 Digitalia

In der deutschsprachigen Geschichtswissenschaft wird das hier behandelte Material etwas sperrig gerne als genuin digital entstandene Quellen bezeichnet.² Ein wenig griffiger hören wir auf Englisch hingegen von den *Born Digital Sources*, so beispielsweise bei Niels Brügger,³ der neben *Digitized* (also digitalisierten analogen Objekten) und *Born-Digital* mit *Reborn* noch eine dritte Kategorie digitalen Materials beschreibt: „Reborn digital material is born-digital material that has been collected and preserved, and that has been changed in this process to such an extent that it is not identical to the born-digital material

1 Vgl. Statista, „Spam-Mails – Anteil am gesamten E-Mail-Verkehr weltweit bis 2019,“ Zugriff am 19.08.2021, <https://de.statista.com/statistik/daten/studie/872986/umfrage/anteil-der-spam-mails-am-gesamten-e-mail-verkehr-weltweit>.

2 So zum Beispiel Catherina Schreiber, „Genuine Internetdaten als historische Quellen – Entwurf einer korrealistischen Quellentheorie,“ *Zeitschrift für digitale Geschichtswissenschaften* 0/1 (2012), <http://hdl.handle.net/10993/7981>.

3 Vgl. Niels Brügger, *The Archived Web: Doing History in the Digital Age* (Cambridge, Massachusetts/London: The MIT Press, 2018)

from which it was made.“⁴ Die Veränderung von digitalen Daten durch Archivierung beziehungsweise Wiederherstellung wird uns im Zusammenhang mit der Überlieferungsgeschichte des Usenet zu einem späteren Zeitpunkt noch einmal begegnen. In Hinsicht auf eine allgemein praktikable Begrifflichkeit möchte ich jedoch für eine noch stärkere Vereinfachung plädieren und deshalb für Objekte, die auf der rein digitalen Ebene entstanden, gespeichert und transportiert worden sind, den Terminus der *digitalia* etablieren. Dazu zählen unter anderem E-Mails, Webseiten, digitale Textdokumente, Social Media-Inhalte, digitale Bilder und Videos, aber eben auch solche Datenbündel wie Logfiles, Programmcode, Bewegungsdaten oder Computerspiele. All diese *digitalia* können bereits für heutige und perspektivisch noch stärker für künftige Historiker:innen von entscheidender Bedeutung sein. Zeitgeschichtliche Forschung ist – zumindest für Themenbereiche, die ab den späten 1980er Jahren angesiedelt sind – ohne die Berücksichtigung dieser Quellen kaum mehr vorstellbar. Für aktuellere Zeiträume stellen sie bereits jetzt manchmal nur das einzig verfügbare Quellenmaterial dar.⁵

Die in Binärcode gespeicherten digitalen Daten repräsentieren nicht direkt die eigentlichen Inhalte, sondern letztlich nur die Codierung dieser Inhalte in maschinenlesbaren Befehlen, die den Computer beispielsweise dahingehend steuern, die hinterlegten Werte als Pixel auf dem Schirm zu zeigen, sie als Buchstaben auf Papier zu drucken oder sie anhand eines programmierten Algorithmus zu verändern. Das bedeutet, dass jeglicher Zugang zu den Daten und ihrer Bedeutung eine maschinelle Decodierung benötigt: „[D]er Umgang mit Information [wird][...] nur durch die Vermittlung von Maschinen rezipierbar.“⁶ Sowohl die Darstellung als auch die Analyse dieses Materials muss also durch eine Übersetzungsleistung des Computers stattfinden. Wir können hierbei in Anlehnung an den etablierten Begriff der *Computer Mediated Communication* fast schon von einer *Computer Mediated History* sprechen.

Oftmals wissen oder verstehen wir dabei nicht genau, wie die Daten durch die angewandten Prozesse verändert oder gefiltert wurden oder noch werden. Umso wichtiger ist es, zum einen die technischen Kontexte der Entstehung, Codierung, Archivierung und Überlieferung dieser Daten zu berücksichtigen und zum anderen die Zusammenhänge bei ihrer Darstellung und Analyse im Sinne

4 Ibidem, 22.

5 Dazu ausführlich Ian Milligan, *History in the Age of Abundance? How the Web Is Transforming Historical Research* (Montreal: McGill Queen's University Press, 2019).

6 Peter Haber, *Digital Past – Geschichtswissenschaft im Digitalen Zeitalter* (München: Oldenbourg Verlag, 2011), 102.

einer Toolkritik im Blick zu behalten.⁷ Diese teilweise komplexen Prozesse der Überlieferung digitaler Daten führen unter anderem zu der paradoxen Situation, dass zwar einerseits ein Überfluss an Informationen vorhanden zu sein scheint, andererseits jedoch größere Bestände aufgrund des Verlustes von passender Software und Hardware nicht mehr benutzbar sind.⁸

Dabei ergibt sich vor allem bei der Beschäftigung mit großen Datenbeständen aus dem Bereich der digitalen Kommunikation – beispielsweise E-Mail, Social Media oder Chatforen – ein für die historische Zunft in diesem Ausmaß ungewohntes Phänomen, das wohl am besten mit dem Begriff *Noise* bezeichnet werden kann. Dieser Kommunikationslärm tritt beispielsweise in Form von unerwünschten Werbenachrichten auf, kann aber auch aus Off-Topic-Diskussionen bestehen oder gar aus gezielten Verbalattacken in Form von so genanntem *Flaming* oder *Trolling*. Diese Störgeräusche können in unterschiedlicher medialer Art aufscheinen. Im Folgenden werde ich mich jedoch auf die textuellen Erscheinungsformen konzentrieren.

Bei einer direkten Beschäftigung mit dem Material im Sinne eines *Close Reading* werden solche Störungen zwar meist relativ problemlos erkannt und können entsprechend kontextualisiert und gegebenenfalls ausgefiltert werden. Gehen wir jedoch ins *Distant Reading* über, lassen also größere Textmengen computergestützt analysieren, kommt es zu einer deutlichen Beeinflussung der Ergebnisse durch den Kommunikationslärm, sei es bei einer simplen Wortzählung, einer Sentimentanalyse oder beim Topic Modeling.⁹ Das ist nicht per se negativ. Für manche Fragestellung kann gerade der Lärm nützlich sein und auf Entscheidendes hinweisen, meist verfälscht er jedoch die erhoffte inhaltliche Erfassung des Materials. Entsprechend muss dieser Effekt, wenn er nicht durch vorgeschaltetes *Data Cleaning* minimiert werden soll (und kann), zumindest einkalkuliert werden.

⁷ Vgl. Andreas Fickers, „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“, *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 157–168, doi: 10.14765/ZZF.DOK-1765.

⁸ Vgl. den Beitrag von Moritz Feichtinger in diesem Band.

⁹ Zum grundsätzlichen Erkenntniswert von Topic Modeling in der historischen Quellenanalyse vgl. Martin Fechner und Andreas Weiß, „Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts“, *Zeitschrift für digitale Geisteswissenschaften* 2/2 (2017), doi: 10.17175/2017_005. Siehe auch den Beitrag von Melanie Althage in diesem Band.

2 Spam im Usenet

Um sowohl die (forschungs-)konzeptionelle als auch die praktische Problematik besser zu veranschaulichen, sollen nun verschiedene Aspekte der Arbeit mit Usenet-Nachrichten der 1990er Jahre aufgegriffen werden, die ab einem gewissen Zeitpunkt immer stärker von Spam betroffen waren.

Es handelt sich beim Usenet um ein Netzwerk von Newsservern, durch das eine textbasierte Kommunikation der Nutzer:innen ermöglicht wird. Nachrichten werden im Usenet in eine hierarchische Struktur unterschiedlicher Newsgroups eingestellt und dezentral gespeichert. Im Usenet entwickelten sich grundlegende Erscheinungsformen der Online-Kommunikation wie die Netiquette oder auch die Zusammenstellung von Frequently Asked Questions (FAQ) zur Regelung des Informationsaustausches, aber auch negative Aspekte wie Flaming oder der Versand von Spam lassen sich dort früh identifizieren. Entwickelt wurde das grundlegende Konzept im Herbst 1979. Im Sommer 1980 war das Projekt ausgereift genug, um auf einer Entwicklerkonferenz in Delaware vorgestellt zu werden. Das UNIX User Network etablierte sich schnell als eine Art „ARPANET für arme Leute“¹⁰. Es war zu Beginn ganz bewusst nach kooperativen Grundprinzipien organisiert, bald gab es jedoch zunehmende Auseinandersetzungen über Meinungsfreiheit und Zensurversuche im Usenet, die 1986 zur Neuorganisation und 1988 letztlich zur Abspaltung der (inhaltlich unabhängigen) Alternative-Hierarchie führten.¹¹

Spam-Nachrichten sind uns aus unseren Mailpostfächern wohlbekannt. Die massenhaft und normalerweise ungerichtet versendeten Nachrichten oder E-Mails haben oft einen kommerziellen Hintergrund, lassen sich also meist als unerwünschte Werbung klassifizieren. Solche Phänomene treten bereits in der Computerkommunikation der 1970er Jahre erstmals auf, und auch wenn die Bezeichnung Spam noch nicht benutzt wird, ist die Motivation von „Spammern“ in dieser Zeit nicht vorwiegend auf kommerzielle Ziele ausgerichtet: „[W]hatever you were posting, be it duplicated, way too long, saturated with quotes,

10 Das ARPANET war die erste systematische Vernetzung von Computern in den USA, die jedoch nur eine enge Auswahl von Computern aus Regierungsinstitutionen und Eliteuniversitäten einschloss. Eine faktisch solide Darstellung der Frühgeschichte des Usenet bieten Michael Hauben und Ronda Hauben, *Netizens: On the History and Impact of Usenet and the Internet* (Los Alamitos, California: IEEE Computer Society Press, 1997). In ihren Bewertungen bleiben die Haubens jedoch oftmals auffällig unkritisch, was eventuell an ihrer persönlichen Nähe zur Usenet Community gelegen haben könnte.

11 Vgl. dazu ausführlicher Oliver Kiechle, „Ein Gespaltenes Netz? – Das Usenet der 1980er-Jahre zwischen Regulierung und Anarchie,“ in *Zur Geschichte des digitalen Zeitalters*, hg. v. Ricky Wichum und Daniela Zetti (Wiesbaden: VS Verlag für Sozialwissenschaften 2022), 125–142.

contextually inappropriate, had broken the implicature of network conversation that held that you should be in some way relevant.“¹² Spam dient also zu Beginn vor allem dazu, die Möglichkeiten technischer Computernetze auszunutzen, um Aufmerksamkeit zu erregen: „Spammers take the infrastructure of the ‚good things‘ and push them to extremes.“¹³

Diese Motivationslage änderte sich mit der Zeit zunehmend in eine kommerzielle Richtung, was unter anderem durch Vorkommnisse im Usenet vorangetrieben wurde. Auch dort, wo man sich zwar grundsätzlich anti-kommerziell gab, können wir schon früh eine Vorstufe des Spammings im Extensive Cross-Posting feststellen. Hier zirkulierten beispielsweise ab und an Angebote von Computerfirmen durch das verpönte gleichzeitige Posten in mehreren Newsgroups. Für Aufsehen und Diskussionen sorgte dann jedoch vor allem ein Aufruf des Nutzers JJ im Jahr 1988, der an Dutzende Newsgroups die Aufforderung richtete, ihn als ziemlich abgebrannten Studenten doch finanziell zu unterstützen.¹⁴ Wurde der JJ-Vorfall innerhalb des Usenet 1988 zwar kurz und heftig in Hinblick auf die inhaltliche Stabilität des Netzwerks und vor allem auf mögliche Sanktionen gegen JJ diskutiert, kehrte schon kurz danach wieder relative Ruhe ein: „Usenet returned to trading anecdotes, sharing knowledge, joking, and bickering, the moment of self-reflexive panic slipped into memory.“¹⁵

Die Bezeichnung Spam wurde in Anlehnung an einen alten Monty Python-Sketch dann Anfang der 1990er ebenfalls im Usenet geprägt. Einhergehend mit dem *Eternal September* des Jahres 1994,¹⁶ als der Internet-Anbieter AOL seinen Kund:innen neben dem einfachen Zugang zum gerade aufstrebenden World Wide Web auch Eingang zum Usenet verschaffte, nahm nicht nur die Zahl der

12 Finn Brunton, *Spam: A Shadow History of the Internet* (Cambridge, Massachusetts: The MIT Press, 2013), 48.

13 Ibidem, 11.

14 Brad Templeton, „Origin of the Term ‚Spam‘ to Mean Net Abuse“ (o.D.), in *Brad Templeton's Home Page, Essays on Junk E-mail (Spam)*. Zugriff am 22.08. 2021, <https://www.templetons.com/brad/spamterm.html>.

15 Brunton, *Spam*, 48.

16 Dieser begann gar nicht im Herbst, sondern bereits im März 1994. Die Bezeichnung rührte daher, dass es im September durch den Beginn des amerikanischen Studienjahres und entsprechend vieler neuer Studierender alljährlich auch eine im Vergleich zum Rest des Jahres höhere Zahl an neuen Usenet-Nutzer:innen gegeben hatte. Nun stieg die Zahl durch das AOL-Angebot aber permanent und deutlich und vor allem über die ursprünglichen universitären (und meist eher liberal eingestellten) Kreise hinaus an, was viele „alte“ Nutzer:innen zunächst irritierte und später sogar verprellte (vgl. Bryan Pfaffenberger, „A Standing Wave in the Web of Our Communication: Usenet and the Socio-Technical Construction of Cyberspace Values,“ in *From Usenet to CoWebs: Interacting with Social Information Spaces*, hg. v. Christopher Lueg und Danyel Fisher (London: Springer, 2003), 40–41).

Usenet-Nutzer:innen rapide zu, sondern die unerwünschte Kommunikation geriet ab Mitte der 1990er Jahre zu einer Massenerscheinung: Große aufsehenerregende Aktionen gab es 1994 zunächst mit der Nachricht „Global Alert for All: Jesus is Coming Soon“, die von einem Studenten an über 200 Newsgroups versendet wurde, und schließlich vor allem durch die Werbung der Anwaltskanzlei Canter und Siegel, die ihr Angebot für Unterstützung bei der Green Card Lotterie automatisiert an fast alle damaligen Newsgroups verschickte.¹⁷ In der Folge wuchs das Spam-Aufkommen im Usenet immens, so dass manche Newsgroups zu Beginn der 2000er Jahre regelrecht geflutet wurden und die eigentlichen Diskussionsthemen komplett untergingen.

Anders als beim JJ-Vorfall reagierte die Community auf Canter/Siegel umgehend, vor allem der automatisierte Versand der Spam-Nachricht hatte viele Systemadministrator:innen aufgeschreckt. Eine erste direkte Reaktion war eine Flut von E-Mails an den Internetanbieter der Kanzlei, der die Anwälte daraufhin prompt vom Netz nahm (eine eher symbolische Aktion, da diese einen Tag später wieder online waren).¹⁸ Weitreichender war jedoch die Reaktion eines bis heute nicht genauer identifizierten Users namens *Cancelmoose*, der einen *Cancelbot* entwickelte. Dieses Programm nutzte im Prinzip den gleichen technischen Ansatz wie die ursprüngliche Spam-Nachricht, wanderte automatisiert durch die Usenet-Newsgroups und löschte die Green Card-Nachricht, wo immer sie zu finden war. Auch wenn diese Vorgehensweise bei den meisten Nutzer:innen Beifall fand, flammten nun erneute Diskussionen über Zensur und Meinungsfreiheit im Usenet auf, zumal Canter und Siegel im Nachgang versuchten, sich als „icons of free speech on Usenet and the Internet generally“ zu gerieren.¹⁹ Um eine Nachricht als Spam klassifizieren und dann gegebenenfalls löschen lassen zu können, akzeptierte die Usenet Community schließlich den nach seinem Erfinder benannten Breidbart-Index als maßgebliche Kennzahl für exzessives Cross-Posting.²⁰

3 Überlieferung von Spam im Usenet

Das Usenet definierte sich von Beginn an als flüchtiges Medium, da vor allem aufgrund der damaligen Limitierungen bei Übertragungsbandbreite und Spei-

¹⁷ Brunton, *Spam*, 48.

¹⁸ Pfaffenberger, „Usenet and the Socio-Technical Construction of Cyberspace Values,“ 36.

¹⁹ Brunton, *Spam*, 57.

²⁰ Pfaffenberger, „Usenet and the Socio-Technical Construction of Cyberspace Values,“ 39.

cherplatzkosten eine längere Speicherung der Nachrichten nicht vorgesehen und kaum möglich war. Entsprechend gibt es erhebliche Lücken in der historischen Überlieferung des Materials, insbesondere bis in die Mitte der 1990er Jahre. Mehrere kleinere Sammlungen und ein größerer Korpus (die so genannten UTZOO-Tapes) alter Usenet-Nachrichten stellen die zentrale Überlieferung von Usenet-Inhalten aus den 1980er und frühen 1990er Jahren dar. Diese wurden im Jahr 2001 zusammen mit den von Google aufgekauften Beständen der Firma *Deja News*, die ab Mitte der 1990er Jahre versucht hatte, das Usenet systematisch zu archivieren und ihren Nutzer:innen zur Verfügung zu stellen, in die neu geschaffenen *Google Groups* integriert.²¹ Theoretisch lassen sich über *Google Groups* nach wie vor eine enorme Anzahl von historischen Usenet-Nachrichten finden, jedoch wurden in den letzten Jahren nicht nur die Suchmöglichkeiten stark eingegrenzt, sondern auch automatisierte Crawl-Prozesse zur individuellen Zusammenstellung eines Korpus weitestgehend geblockt. Anhand der *Historical Usenet Collection* des *Internet Archive* zeigt sich jedoch, dass diese Blockade offenbar umgangen werden kann.²²

Analog zur gängigen Archivpraxis bedeutet Webarchivierung, dass Quellen selektiert, gesammelt und geordnet werden müssen. Bei *digitalia* finden meist zusätzlich technische Transformationsprozesse statt: Für die Archivierung müssen Daten eventuell bereinigt und gefiltert oder aber auch mit zusätzlichen Metadaten versehen werden, bevor sie in ein Archivformat wie zum Beispiel WebARCHive (WARC) für Internetinhalte überführt werden. Eine bei Usenet-Inhalten zusätzliche Problematik stellt der seit der Veröffentlichung von archivierten Usenet-Nachrichten durch *Deja News* und einer darauf folgenden massiven Gegenwehr eines Teils der Community aufgrund von Datenschutzbedenken eingeführten X-No-Archive-Header dar. Diese Einstellung im Header der Nachricht verhindert im Normalfall eine Archivierung durch automatisierte Prozesse. Zusätzlich besteht auch nach wie vor die Möglichkeit, eigene Postings auf Antrag aus der Archivierung entfernen zu lassen.

Für Spam-Nachrichten aus dem Usenet bedeutet das vor allem, dass diese zum Teil in die archivierten Usenet-Bestände übernommen wurden, zum Teil aber eben auch nicht: Entweder wurden sie bereits vor dem Archivieren systematisch gelöscht (durch Cancelbots) beziehungsweise durch die jeweiligen Nutzer:innen lokal ausgefiltert oder dann während des Archivierungsprozesses ent-

21 Camille Paloque-Bergès, „Usenet as a Web Archive: Multi-Layered Archives of Computer-Mediated-Communication,“ in *Web 25: Histories from the First 25 Years of the World Wide Web*, hg. v. Niels Brügger (New York et al., Peter Lang Publishing, 2017), HAL:halshs-01843627, 236.

22 Zur Usenet-Archivierung vgl. Kiechle, „Das Usenet der 1980er-Jahre zwischen Regulierung und Anarchie“, 127–130.

fernt. Nichtsdestotrotz finden sich in einigen archivierten Newsgroups nach wie vor eine große Anzahl an Spam-Nachrichten, was unter anderem daraus resultiert, dass Spam in *Google Groups* grundsätzlich nicht pauschal ausgefiltert wird.

Ein weiteres Problem von öffentlich zugänglich gemachten Usenet-Nachrichten besteht darin, dass Nutzer:innen im Vertrauen auf die angenommene Flüchtigkeit der Kommunikation dort oftmals ihre E-Mail-Adressen angegeben haben. Diese können nun automatisiert eingesammelt und zum Versand weiterer Spam-Nachrichten benutzt werden. Diesem Umstand trägt beispielsweise das erst seit 2021 existierende Portal *usenetarchives.com* dadurch Rechnung, dass persönliche Informationen auf den zur Verfügung gestellten historischen Usenet-Nachrichten weitgehend unkenntlich gemacht wurden.²³

4 Flaming und Bots

Neben Spam lassen sich noch weitere Arten von Lärm in *digitalia* finden, die an dieser Stelle kurz beschrieben werden sollen: Das *Flaming* beziehungsweise *Trolling* und so genannte *Social Bots*.

Das Phänomen des Flaming, also abwertende oder gar beleidigende Einlassungen, lässt sich schon bei den ersten Gehversuchen der Online-Kommunikation in den 1970er Jahren erkennen, massiv tritt es dann mit der Etablierung und dem raschen Anwachsen des Usenet auf. Beim Flaming handelt es sich um das bewusste Versenden von provozierenden oder beleidigenden Nachrichten, beispielsweise in Online-Foren, Bulletin Boards oder Newsgroups. Die wortnahe Übersetzung von Flaming mit „zündeln“ passt zu diesem Vorgang ganz gut. Es geht meist darum, sachliche Diskussionen zu stören oder gar zu sprengen. Nicht selten entzündet sich an gleichgelagerten Reaktionen auf einen Flame ein regelrechter *Flame War*, der die ursprüngliche Intention einer Online-Unterhaltung komplett in den Hintergrund rückt.²⁴

²³ Usenet Archives, <https://www.usenetarchives.com/>, online seit Mai 2021. Hier lassen sich archivierte Usenet-Nachrichten relativ komfortabel nach inhaltlichen Aspekten durchsuchen. Jedoch ist zum einen nicht ganz klar, welche historischen Usenet-Bestände dort überhaupt eingebunden sind und zum anderen erschwert das Fehlen persönlicher Informationen – so nachvollziehbar diese Herangehensweise in Bezug auf Datenschutzaspekte auch ist – teilweise eine tiefer gehende Analyse.

²⁴ Michael Nycyk, „Trolls and Trolling History From Subculture to Mainstream Practices,“ in *The SAGE Handbook of Web History*, hg. v. Niels Brügger und Ian Milligan (Los Angeles et al.: SAGE, 2019), 580.

Im Usenet wurde früh versucht, dieses Verhalten einzudämmen oder zumindest zu kanalisieren. So wurde bereits in den ersten Versionen der als Netiquette bekannt gewordenen Verhaltensregeln für das Posten im Usenet das Flaming mehr oder weniger geächtet.²⁵ Da es aber zum einen kaum Sanktionsmöglichkeiten gegen diejenigen gab, die Flames posteten und zum anderen viele Nutzer:innen Gefallen an gelungenen Beleidigungen fanden, führte dieser Schritt meist ins Leere. Bei manchen Usern, die nur gelegentlich und oft eher sarkastisch gemeint ausfällig werden wollten, bürgerte es sich ein, den Flame mit den Tags *Flame On/Flame Off* zu kennzeichnen und dadurch als unsachlichen Einwurf sichtbar zu machen.²⁶ Eine Möglichkeit, Flames zu kanalisieren, war die Einrichtung der Newsgroup *net.flames*, auf die aus dem Ruder laufende Diskussionen abgeleitet werden sollten, was jedoch nur zum Teil gelang.

Im Usenet der frühen 1980er war es üblich, seinen echten Namen zu benutzen und meist sogar noch die postalische Adresse anzugeben, was justiziable Beschimpfungen eher selten machte. Die zunehmende Anonymisierung und vor allem eine immens steigende Nutzer:innenzahl führte zu immer schlimmeren Flames. Ab den 1990ern findet sich das Phänomen dann in Form des Trollings auf beinahe allen Kommunikationsebenen des World Wide Web wieder.²⁷ In den heutigen Social Media sind solche Ausfälle weit verbreitet und häufig mit Begriffen wie Hate-Speech, Shitstorm oder Fake News verknüpft.

Ab den 2000er Jahren machen sich zudem vermehrt Social Bots in der Online-Kommunikation bemerkbar, obwohl auch diese beileibe keine neue Erscheinung waren. Bei Bots handelt es sich um algorithmisch gesteuerte und oft mit Hilfe von Machine-Learning programmierte Tools, die versuchen, die menschliche Kommunikation im Internet im Sinne ihrer Programmierer:innen

25 Vgl. eine der ersten und vermutlich am meisten rezipierte Kodifizierung der Usenet-Netiquette: Chuq v. Rospach, „A Primer on How to Work with the USENET Community,“ Zugriff am 05.07.2019, <http://www.mit.edu/afs.new/athena/activity/i/isig/Primer/usenet-primer>.

26 Denis Howe, „Flame from FOLDOC,“ *Free Online Dictionary of Computing*, Zugriff am 30.08.2021, <https://foldoc.org/flame>.

27 Die Bezeichnung „Troll“ wurde offenbar ebenfalls im Usenet geprägt, vgl. Nycyk, „Trolls and Trolling History,“ 580.

zu simulieren und zu beeinflussen.²⁸ Wir finden diese Programme schon im Usenet der 1990er Jahre beispielsweise in Form der vorher erwähnten Cancelbots. Dabei handelt es sich aber meist um Einzelaktionen. Mit Auftreten der großen Social Media Plattformen in den 2000er Jahren halten Social Bots relativ schnell massenhaften Einzug in die Online-Kommunikation. Diese agieren meist zweckgerichtet: Es geht zum Beispiel darum, bestimmte Kanäle mit auf den ersten Blick regulären Messages zu blockieren, personalisierte Werbung für bestimmte Produkte zu platzieren, Trends zu kreieren oder schlicht um Trolling. In diesen Zusammenhang sind aktuelle Diskussionen um eine mögliche Beeinflussung politischer Wahlen durch Bots in sozialen Netzwerken einzuordnen.

5 *Noisy digitalia* in der historischen Analyse

Wie sollten wir nun mit diesen Stördaten in unseren digitalen Quellen, seien es Flame Wars oder Hate Speech, Spam-Fluten oder nichtmenschliche Kommunikation, in der historischen Arbeit umgehen? Zwei grundsätzliche Möglichkeiten scheinen sich mir hierbei anzubieten: Entweder versuchen, sie einzubinden mit all den Weiterungen, die dies nach sich ziehen könnte oder den Lärm weitestgehend auszufiltern, was aber Folgen für die Validität der Daten bzw. der Inhalte haben kann.

Wie eingangs erwähnt, erscheint es wenig problematisch, wenn wir das verwendete Material selbst ansehen und den *Noise* direkt erkennen und einordnen können; nichts anderes beinhaltet die Arbeit mit nicht-digitalen schriftlichen Quellen. Arbeiten wir jedoch mit Hilfe des Computers und analysieren unsere *digitalia* in größerem Maßstab, können umfangreiche Vorarbeiten im Sinne eines Data Cleaning und Preprocessing nötig werden. Bezogen auf die vorher ausgeführten Störkategorien kann sich dies unterschiedlich problematisch gestalten.

In Bezug auf Spam-Nachrichten funktioniert eine Filterung verhältnismäßig gut. Die gewohnten Spam-Filter orientieren sich an bestimmten Schlüsselwörtern, um entsprechende Nachrichten auszufiltern und lassen sich zum Teil im Nachhinein anwenden. Dies geht sowohl mit E-Mails als auch mit Usenet-Nachrichten, da beide Formate zum einen ursprünglich auf demselben RFC-Standard 822 („Standard for the Format of ARPA Internet Text Messages“) beruhen und

²⁸ Vgl. Joseph Seering et al., „The Social Roles of Bots. Evaluating Impact of Bots on Discussions in Online Communities,“ *Proceedings of the ACM on Human-Computer Interaction* 2, Computer Supported Cooperative Work (CSCW) (2018): 1, doi: 10.1145/3274426.

zum anderen manche Mailprogramme zugleich als Newsreader fungieren und entsprechend die gleichen Filter-Plugins angewandt werden können. Trotzdem muss jeweils nachgehalten werden, ob die Filteroptionen nicht auch normale Nachrichten aussortieren. Bei aktuellen historischen Arbeiten zum Usenet wurde das Spam-Problem adressiert: Avery Dame-Griff beschreibt bei seiner Untersuchung von Transgender-Newsgroups das zunehmende Auftreten von Spam-Nachrichten, die nicht nur Werbebotschaften, sondern auch automatisierte Anti-Transgender-Postings transportierten.²⁹ Aus diesem Grund entschied er sich, diese Nachrichten nicht auszufiltern, „as they formed an inextricable part of the Usenet experience. Nevertheless, this choice did consequently limit the applicability of more linguistically focused methodology, such as topic modelling.“³⁰ In ihrer Doktorarbeit zur LGBTQ-Identität im frühen Internet beschäftigt sich Sarah McTavish ebenfalls mit entsprechenden Newsgroups und stellte in Hinblick auf die Gruppe *alt.sex.motss* fest: „[T]he group quickly became consumed with spam messages, both explicit and not.“³¹ Inwiefern diese Nachrichten Teil ihrer Analyse sind, wird jedoch leider offen gelassen.

Wenn wir den Blick auf das Flaming wenden, gibt es bereits Ende der 1990er Jahre, als die Sozialwissenschaften speziell das Usenet als Untersuchungsgegenstand für sich entdecken, erste Versuche, Tools zu entwickeln, die Flames erkennen und ausfiltern können.³² Grundsätzlich arbeiten diese Anwendungen wie moderne Hate-Speech-Detektoren, mit Machine-Learning-Methoden. Das Programm erlernt also bestimmte Schlüsselbegriffe und Kommunikationsmuster und versucht damit, entsprechende Inhalte zu erkennen.³³ Abgesehen davon, dass der Erfolg dieser Methoden wie immer von den Trainingsdaten abhängt, findet Flaming – das sich oft ironisch oder sarkastisch darstellt – manchmal so subtil statt, dass diese Tools an ihre Grenzen stoßen.

Die technischen Möglichkeiten, Bots *ex post* zu identifizieren, sind ebenfalls begrenzt, da die entsprechenden Programme meist versuchen, den ursprüngli-

29 Vgl. Avery Dame-Griff, „Herdning the ‚Performing Elephants:‘ Using Computational Methods to Study Usenet,“ *Internet Histories* 11/10 (2019): 12, doi: 10.1080/24701475.2019.1652456.

30 Ibidem.

31 Sarah McTavish, „The Queer Eternal September: LGBTQ Identity on the Early Internet and Web“ (Dissertation University of Waterloo, Ontario, 2020), URI: <http://hdl.handle.net/10012/16336>, 105.

32 Melanie Martin, *Annotating Flames in Usenet Newsgroups: A Corpus Study* (NSF Minority Institution Infrastructure Grant Site Visit to NMSU CS Department, 2002). Zugriff am 30.08.2021, https://www.cs.nmsu.edu/~mmartin/pubs/martin_poster.pdf.

33 Vgl. Sasha Sax, *Flame Wars: Automatic Insult Detection* (Stanford University, 2016), Department of Computer Science Report, Stanford University, Zugriff am 30.08.2021, <http://cs224d.stanford.edu/reports/Sax.pdf>.

chen Absender per IP-Adresse zurückzuverfolgen, um einen Bot erkennen zu können. Dies ist bei archivierten *digitalia* jedoch nicht mehr ohne weiteres möglich. Auch imitieren moderne Bots das menschliche Kommunikationsverhalten oft so echt, dass es kaum mehr von dem echter Nutzer:innen zu unterscheiden ist.³⁴ Das führt zu der fast schon paradoxen Situation, dass künstliche Intelligenz – wenn wir fortgeschrittene Machine-Learning-Technik so nennen wollen – versucht zu erkennen, ob eine andere künstliche Intelligenz die Kommunikation von Menschen imitiert hat.

Grundsätzlich würde ich die Frage nach Einbinden oder Ausfiltern von der jeweiligen Forschungsperspektive abhängig machen wollen. Wenn durch Distant Reading alle Aspekte eines digitalen Kommunikationsclusters erfasst werden sollen, verbietet sich eine Filterung fast von selbst. Auch können solche negative Erscheinungen zu entscheidenden Hinweisen auf Bereiche werden, die Close Reading erfordern. Ebenfalls sollte nicht unterschätzt werden, dass Spam und noch mehr das Flaming ein natürlicher Teil der historischen Internetkommunikation ist, der auch nonkonforme Akteur:innen – Rebell:innen, Brigant:innen, Anarchist:innen – repräsentiert.

Geht es jedoch um klar eingegrenzte Inhalte, die vom Noise offensichtlich massiv gestört werden, ist eine starke Filterung sicherlich das Mittel der Wahl, nicht zuletzt um eine sinnvolle linguistische Analyse zu ermöglichen. Grundsätzlich besteht jedoch bei allen Verfahren des Data Cleanings die Gefahr einer determinierenden Vorgehensweise. Die Parameter der Filter werden dann, überspitzt gesagt, einfach so lange angepasst, bis die Daten zur vorgegebenen Fragestellung passen. Dadurch besteht, wie durch die unkritische Verwendung von Tools, deren genaue Funktionsweise nicht nachvollziehbar ist, letztlich die Gefahr, das vielbeschworene „Vetorecht“ der Quellen zu unterminieren, also falsche, beziehungsweise nicht mehr durch das Quellenmaterial gedeckte Deutungen zu provozieren.

In diesem Zusammenhang soll noch einmal betont werden, dass die Kenntnis davon, unter welchen Prämissen bestimmte *digitalia* archiviert worden sind, entscheidend bei der Einordnung der gespeicherten Daten ist. Bei der Webarchivierung wird beispielsweise nicht alles grundsätzlich gespeichert, sondern sowohl aufgrund technischer Kapazitäten, als auch aus Relevanzgründen im Vorfeld die speicherwürdigen Webseiten ausgewählt, mit der Gefahr, bestimmte Randaspekte auszuklammern und nicht zu überliefern. In den letzten Jahren hat eine engagierte Diskussion über die ethischen Aspekte der Webarchivierung

³⁴ Vgl. Adrian Rauchfleisch und Jonas Kaiser, „The False Positive Problem of Automatic Bot Detection in Social Science Research,“ *PLOS ONE* 15/10 (2020), doi: 10.1371/journal.pone.0241045.

zudem dazu geführt, dass manche Inhalte ganz bewusst nicht archiviert oder massiv gefiltert werden, um beispielsweise Opfer rassistischer oder homophober Attacken im Netz zu schützen.³⁵ Dies führt ganz allgemein zu dem Problem, dass hierbei viele rechtliche Fragen, namentlich bezüglich der Urheber- und Persönlichkeitsrechte, noch nicht ausreichend geklärt sind. Diese Probleme betreffen zum einen den Archivierungsprozess, zum anderen aber auch massiv die Forschungsarbeit mit rechtlich heiklen Beständen. Zwar kann die Analyse und Bewertung des Materials eventuell in einem rechtlich sicheren Rahmen stattfinden, soll jedoch im Sinne eines guten Forschungsdatenmanagements die Datengrundlage nachverfügbar gemacht werden, könnte dies problematisch werden. Die Arbeit mit *digitalia* hält jedenfalls offensichtlich noch einige Unwägbarkeiten bereit.

Bibliographie

- Brügger, Niels. *The Archived Web: Doing History in the Digital Age*. Cambridge, Massachusetts, London, England: The MIT Press, 2018.
- Brunton, Finn. *Spam: A Shadow History of the Internet*. Cambridge, Massachusetts: The MIT Press, 2013.
- Dame-Griff, Avery. „Herding the ‚Performing Elephants:‘ Using Computational Methods to Study Usenet.“ *Internet Histories* 11/10 (2019): 1–22, doi: 10.1080/24701475.2019.1652456.
- Fechner, Martin und Andreas Weiß. „Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts.“ *Zeitschrift für digitale Geisteswissenschaften* 2/2 (2017), doi: 10.17175/2017_005.
- Fickers, Andreas. „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“ *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 157–168, doi: 10.14765/ZZF.DOK-1765.
- Graham, Pamela M. „Guest Editorial: Reflections on the Ethics of Web Archiving.“ *Journal of Archival Organization* 14/3–4 (2017): 103–110. doi: 10.1080/15332748.2018.1517589.
- Haber, Peter. *Digital Past – Geschichtswissenschaft im Digitalen Zeitalter*, München: Oldenbourg Verlag, 2011.
- Hauben, Michael und Ronda Hauben. *Netizens: On the History and Impact of Usenet and the Internet*. Los Alamitos, California: IEEE Computer Society Press, 1997. Zugriff am 30.08.2021, <http://www.loc.gov/catdir/bios/wiley046/97005920.html>.
- Howe, Denis. „Flame from FOLDOC.“ *Free Online Dictionary of Computing*, Zugriff am 30.08.2021, <https://foldoc.org/flame>.

35 Pamela M. Graham, „Guest Editorial: Reflections on the Ethics of Web Archiving,“ *Journal of Archival Organization* 14, 3–4 (2017): 103–110, doi: 10.1080/15332748.2018.1517589.

- Kiechle, Oliver. „Ein gespaltenes Netz? – Das Usenet der 1980er-Jahre zwischen Regulierung und Anarchie.“ In *Zur Geschichte des Digitalen Zeitalters*, hg. v. Ricky Wichum und Daniela Zetti, 125–142. Wiesbaden: VS Verlag für Sozialwissenschaften, 2022.
- Martin, Melanie. „Annotating Flames in Usenet Newsgroups: A Corpus Study.“ NSF Minority Institution Infrastructure Grant Site Visit to NMSU CS Department, 2002, Zugriff am 30.08.2021, https://www.cs.nmsu.edu/~mmartin/pubs/martin_poster.pdf.
- McTavish, Sarah. *The Queer Eternal September: LGBTQ Identity on the Early Internet and Web*, Dissertation University of Waterloo, Ontario, 2020, URI: <http://hdl.handle.net/10012/16336>.
- Milligan, Ian. *History in the Age of Abundance? How the Web Is Transforming Historical Research*. Montreal: McGill Queen's University Press, 2019.
- Nycyk, Michael. „Trolls and Trolling History From Subculture to Mainstream Practices.“ In *The SAGE Handbook of Web History*, hg. v. Niels Brügger und Ian Milligan, 577–89. Los Angeles, London, New Delhi, Singapore, Washington, DC, Melbourne: SAGE reference, 2019.
- Paloque-Bergès, Camille. „Usenet as a Web Archive: Multi-Layered Archives of Computer-Mediated-Communication.“ In *Web 25: Histories from the First 25 Years of the World Wide Web*. Hg. v. Niels Brügger. Peter Lang Publishing, 2017. HAL: halshs-01843627.
- Pfaffenberger, Bryan. „A Standing Wave in the Web of Our Communication': Usenet and the Socio-Technical Construction of Cyberspace Values.“ In *From Usenet to CoWebs: Interacting with Social Information Spaces*, hg. v. Christopher Lueg und Danyel Fisher. Computer Supported Cooperative Work. London: Springer, 2003.
- Rauchfleisch, Adrian, und Jonas Kaiser. „The False Positive Problem of Automatic Bot Detection in Social Science Research.“ *PLOS ONE* 15/10 (2020): e0241045. doi: 10.1371/journal.pone.0241045.
- Rospach, Chuq V. „A Primer on How to Work with the USENET Community.“ Zugriff am 05.07.2021, <http://www.mit.edu/afs.new/athena/activity/i/isig/Primer/usenet-primer>.
- Sax, Sasha. „Flame Wars: Automatic Insult Detection.“ Department of Computer Science Report, Stanford University, Zugriff am 30.08.2021, <http://cs224d.stanford.edu/reports/Sax.pdf>.
- Schreiber, Catherina. „Genuine Internetdaten als historische Quellen – Entwurf einer korrealistischen Quellentheorie.“ *Zeitschrift für digitale Geschichtswissenschaften* 0/1 (2012): 1–15. <http://hdl.handle.net/10993/7981>.
- Seering, Joseph, Juan P. Flores, Saiph Savage und Jessica Hammer. „The Social Roles of Bots.“ *Proceedings of the ACM on Human-Computer Interaction* 2, Computer Supported Cooperative Work (CSCW) (2018): 1–29. doi: 10.1145/3274426.
- Statista. „Spam-Mails – Anteil am gesamten E-Mail-Verkehr weltweit bis 2019 | Statista.“ Zugriff am 19.08.2021. <https://de.statista.com/statistik/daten/studie/872986/umfrage/an-teil-der-spam-mails-am-gesamten-e-mail-verkehr-weltweit/>.
- Templeton, Brad. „Origin of the Term ‚Spam‘ to Mean Net Abuse.“ (o.D.), in Brad Templeton's Home Page, Essays on Junk E-mail (Spam). Zugriff am 22.08.2021, <https://www.templetons.com/brad/spamterm.html>.

